

I. THE CLAIMED INVENTION

The claimed invention is directed to a method for searching files stored on a network. The method includes accessing a first file on the network, downloading data from the first file and setting an accessing time to access a second file based on data downloaded from the first file.

Conventional network file search engines conduct searches for updated files on networks periodically, such as at regular intervals. One problem with these conventional systems is that these systems do not have any method for determining when a website might be scheduled to be updated. Depending on how often a website is updated, the web crawler's archive data could be very outdated. On the other hand, frequent web crawler visits to websites which are not frequently updated consumes valuable computer resources.

The present invention provides a method for determining when and how often a web crawler should return to a website. The present invention provides this advantage because the method accesses a first file on a network, downloads data from the first file and sets an access time to access a second file based upon the data from the first file.

II. THE PRIOR ART REJECTIONS

The Examiner alleges that the Meyerzon et al. reference teaches the claimed invention. Applicant submits, however, that there are elements of the claimed invention which are neither taught nor suggested by this reference.

The Meyerzon et al. reference discloses synchronizing crawler with a notification source. The Meyerzon et al. reference is concerned with reducing the interval between the time a

document is revised on a Web server and the time until that document is revisited by a Web crawler (col. 1, lines 63 - 65). This interval creates a latency in the index of the Web crawler (col. 1, lines 65-66). The Meyerzon et al. reference teaches that to address this problem it is desirable to avoid checking all documents in a Web crawler's index at regular intervals by providing a Web crawler which only visits those documents which have been changed (col. 2, lines 6-19). Although, these checks are still performed at regular intervals.

The Meyerzon et al. reference does not teach or suggest the features of the claimed invention including setting an accessing time to access a second file based on data downloaded from a first file.

Rather, the Meyerzon et al. reference discloses providing a Web crawler which includes a gatherer project 206 at a local server 204 and a notification source 250 at a remote server 218. The gatherer project 206 includes a gatherer process 240, a notification log 243, a transaction log 242 and a history map 244. The history map 244 includes a list of all retrieved URLs (col. 9, lines 28 - 30). The transaction log 242 includes a list of URLs for documents to gather which is recursively generated (col. 8, lines 65-67). The notification log 243 includes a list of changed URLs received from the notification source 250.

The history map 244 stores a timestamp for each retrieved document which indicates the last time the document was retrieved by the gatherer (col. 3, lines 37 - 40). This history map is used in initialization crawls to minimize the retrieval of documents that the gatherer process has previously crawled and that have not changed since the previous crawl (col. 5, lines 25 - 28).

Contrary to the allegations of the Office Action, the Meyerzon et al. reference does not set an access time. Indeed, the Meyerzon et al. reference does not teach or suggest setting any time for performing any process. The Meyerzon et al. reference only discloses that during an initialization crawl a timestamp for a document in a history map is compared with the timestamp of the document on the remote server being crawled (step 556 of Fig. 5B). If the timestamp of the document being crawled is later (more recently updated) than the timestamp in the history map then the document is retrieved (step 558). If the timestamp of the document is not later, then the URL is marked as having been processed and the history map is updated with the URL and the timestamp (step 564) (see also col. 14, lines 34 - 65).

In other words, the Meyerzon et al. reference teaches performing conventional periodic web crawls, but conditions the update of any URLs upon the determination that the particular document currently being crawled has been changed since the last crawl. In this manner, the Meyerzon et al. reference minimizes latency by reducing the number of files that are updated in the index (col. 14, lines 54 - 65). The Meyerzon et al. reference does not teach or suggest setting an accessing time to access a second file based on data downloaded from a first file. Therefore, the Examiner is respectfully requested to withdraw this rejection.

In summary, the present invention does not require a notification source which periodically visits a Web site looking for changes as disclosed by the Meyerzon et al. reference. Rather, the present invention relies upon data that already exists on a site to determine when to revisit a site, as a result of setting an access time.

III. FORMAL MATTERS AND CONCLUSION

In view of the foregoing amendments and remarks, and the discussion during the November 13, 2002 telephone interview, Applicant respectfully submits that claims 1-23, all the claims presently pending in the Application, are patentably distinct over the prior art of record and are in condition for allowance. The Examiner is respectfully requested to pass the above application to issue at the earliest possible time.


Should the Examiner find the Application to be other than in condition for allowance, the Examiner is requested to contact the undersigned at the local telephone number listed below to discuss any other changes deemed necessary in a telephonic or personal interview.

The Commissioner is hereby authorized to charge any deficiency in fees or to credit any overpayment in fees to Attorney's Deposit Account No. 50-0481.

Respectfully Submitted,

Date: _____

11/14/02



James E. Howard
Registration No. 39,715

McGinn & Gibb, PLLC
8321 Old Courthouse Rd., Suite 200
Vienna, Virginia 22182
(703) 761-4100
Customer No. 21254